Chapter 7

GROWTH POINTS IN THINKING-FOR-SPEAKING[1]

David McNeill and Susan D. Duncan

1. INTRODUCTION

Many bilingual speakers believe they engage in different forms of thinking when they shift languages. This experience of entering different thought worlds can be explained with the hypothesis that languages induce different forms of 'thinking-for-speaking'— thinking generated, as Slobin (1987) says, because of the requirements of a linguistic code. "'Thinking for speaking' involves picking those characteristics that (a) fit some conceptualization of the event, and (b) are readily encodable in the language"[2] (p. 435). That languages differ in their thinking-for-speaking demands is a version of the linguistic relativity hypothesis, the proposition that language influences thought and that different languages influence thought in different ways.

Thinking-for-speaking differs from the so-called 'strong' Whorfian version of the linguistic relativity hypothesis, as we understand it. The latter (Whorf, 1956; Lucy, 1987, 1992) refers to general, *langue*-wide patterns of 'habitual thought', patterns that, according to the hypothesis, are embodied in the forms of the language and analogies among them. The thinking-for-speaking hypothesis, in contrast, refers to how speakers organize their thinking to meet the demands of linguistic encoding on-line, during acts of speaking — what Saussure termed *parole* rather than *langue* (Saussure, 1959). The thinking-for-speaking version and the Whorfian version of the linguistic relativity hypothesis are not mutually exclusive, but neither are they identical. The distinction between them parallels the characterization of Whorf as 'synchronic' compared to Vygotsky (1987) as 'diachronic' that was offered by Lucy & Wertsch (1987). Following them, we will regard the thinking-for-speaking hypothesis as having a diachronic focus on *thinking*, rather than a synchronic focus on *habitual thought*.

Slobin outlined three approaches to demonstrate linguistic relativity in this thinking-for-speaking sense. One is to find the stages at which children talk about experience in ways that

appear specifically shaped by the linguistic system they are acquiring; another is to identify the difficulties that second language learners have in adapting their thinking to the new language; the third is to look at languages historically — the elements most resistant to change being possibly those most deeply ingrained in thought. In each of these approaches, spoken language is the only source of information, and none of them breaks into the logical circle of language representing itself. Child speakers of different languages acquire expressive habits that mirror the semantic-structural differences among their languages (Choi and Bowerman, 1991). A skeptical view, however, could hold that these differences operate only at the level of linguistic expression. To counter such a view, some way is needed to externalize cognition in addition to language. In this chapter, we consider speech and gesture jointly as an enhanced 'window' onto thinking and show how the co-occurrences of speech and gesture in different languages enable us to infer thinking-for-speaking in Slobin's sense.

2. THE GESTURES WE ANALYZE

2.1. Co-expressive and synchronized with speech

Detailed observations of the gestures that accompany speech show that gesture and speech are systematically organized in relation to one another. The gestures are meaningful. They form meaningful, often nonredundant combinations with the speech segments with which they synchronize. A speaker raises her hand upward to mean that a character in a story is climbing up. The rising hand expresses upwardness, and so does the speech, "[and he climbs **up the pipe**]". The specific phase of the gesture depicting upwardness coincides with the semantically most congruent parts of the utterance (the stroke phase shown with boldface). The confluence of speech and gesture suggests that the speaker was thinking in terms of a *combination* of imagery and linguistic categorial content; 'thought' was in the form of a holistic image of the character rising upward coordinated with the analytic, linguistically categorized meanings of 'up' and 'the pipe'. The contents of the gesture and the synchronized speech need not be identical, and indeed they usually are not. An upward gesture combined with "up" might be regarded as redundant, but the same gesture continued through "the pipe" — a related but distinct concept. The term we use to

denote such related but not identical meanings is 'co-expressive.' This means that the gesture and its synchronized co-expressive speech express the same underlying idea unit but do not necessarily express identical aspects of it. By looking at the speech *and* the gesture, jointly, we are able to infer characteristics of this underlying idea unit that may not be obvious from the speech alone.

2.2. Unique semiotic properties

2.2.1. Idiosyncratic. The gestures we analyze are 'idiosyncratic' in the sense that they are not held to standards of good form; instead they are created locally by speakers while they are speaking.[3] The forms of such gestures are driven almost entirely by meaning. The hand rising upward expresses upwardness; upward at an angle expresses upwardness at an angle. The system is graded, not categorial.   Idiosyncratic gestures differ from arbitrary sound-meaning pairings in which meaning plays no role in determining signifier shape. By virtue of idiosyncrasy, co-expressive, speech-synchronized gestures open a 'window' onto thinking that is otherwise curtained. Such a gesture *displays* mental content, and does so instantaneously, in real-time (as prosody also functions to display content; cf. Bolinger, 1986).   Idiosyncratic gestures should be distinguished from gestures of the kind that Kendon has called 'quotable' and others have called 'emblems' (Kendon, this volume; 1992; Morris, et al. 1979; Ekman and Friesen, 1969) — gestures that *must* be configured according to pre-established standards of form in order for them to function as signs, such as the OK sign among North Americans or the *mano a borsa* or purse-hand among Neapolitans (Kendon, this volume).   The OK sign, made not with the forefinger but with the middle finger touching the thumb, is not recognizable as the same sign nor as a graded approximation to it, even though it differs but minimally from the official sign.

2.2.2. Global and synthetic. The way in which idiosyncratic gestures display their mental content is unlike the sign mechanisms of speech. The terms 'global' and 'synthetic' characterize how gestures differ as semiotic entities.

Global: The meanings of the 'parts' or the features of gestures are determined by the meaning of the whole. The meaning determination is whole-to-part. The global property

contrasts with the 'compositional' property of speech. In speech, whole meanings are built up out of independently meaningful parts (morphemes) and the semiotic direction is part-to-whole. It is not the case that the hand in the 'up' gesture, or its movement or direction, separately from this gesture, meant the character, his movement, or direction. These 'parts' received their meanings because they were the parts of a gesture that meant, as a whole, the character rising up. 'Global' in this sense does not refer to the contextualization of meaning, but to how the meanings of the parts are determined by the meanings of wholes.

Synthetic: Distinguishable meanings in speech converge into one symbolic form, the gesture. The synthetic property contrasts to the analytic distribution of meanings across the surface structures of sentences. Putting the semantic components of an actor, action and path-direction in a sentence, "he climbs up the pipe", spreads them out analytically. The upward rising gesture compresses them into one symbol, synthetically.

Thus, when gesture and speech combine, they bring into one meaning system two distinct semiotic architectures. Each modality, because of its unique semiotic properties, can go beyond the meaning possibilities of the other, and this is the foundation of our use of gesture as an enhanced window into mental processes (cf. Kita, this volume, for a related approach).

3. THE GROWTH POINT

The growth point (GP) is the name we give to an analytic unit combining imagery and linguistic categorial content. We center our analysis on this hypothesized unit (a preliminary version of the GP concept was presented in McNeill, 1992).

3.1. Composition and integrity of growth points

GPs are inferred from the totality of communicative events with special focus on speech-gesture synchrony and co-expressivity. Following Vygotsky (1987), a GP is assumed to be a *minimal* psychological unit; that is, the smallest unit (in his analysis) that retains the essential properties of a whole, in our case the whole of an image and a linguistically-codified meaning category, such as we see in the speech-gesture window. We use the gesture's semantic content

and its synchrony (that is, the synchrony of the gesture stroke phase) with speech to infer the GP. For example, to locate the GP of the following,[4]

(1)   and Tweety Bird runs and gets a bowling b[all and Ø <u>drops</u> **it do**<u>wn the drainpipe]</u>

(where the two hands appear to form a large round object and move it down)[5] we refer to both gesture and speech. The GP was embodied in both the image and the synchronized linguistic categorial content. The image was of a cartoon character dropping something down (a bowling ball). The categorial content was the linguistic segments, "it" and "down". The gesture suggests visuo-spatial/actional thinking in which the downward movement of the ball due to the action of an agent was central. Such imagery is important. It grounds the linguistic categories in a specific visuo-spatial context (the gesture itself is not the 'context'; see Section 3.2). The downward content of the gesture is a specific case of the general linguistic category "down" — a specific visualization of it. The linguistic categorization is also crucial, since it brings the image into the system of categories of the language.

The psycholinguistic reality of the GP is seen in the fact that it strongly resists forces trying to divide it. For example, delayed auditory feedback grossly disrupts speech timing but speech-gesture synchrony remains intact (McNeill 1992, Chapter 10, first DAF experiment). Synchrony is disrupted only if speech and gesture are drained of meaning through repetition; i.e., such that GPs may be circumvented in their production (second DAF experiment). Neither does clinical stuttering interrupt speech-gesture synchrony, despite massive disruptions of speech. Gestures during stuttering bouts freeze into holds (Mayberry & Jaques, this volume; also Nobe, this volume). On the reception side, listeners, after a brief delay, cannot tell whether information was conveyed in gesture or in speech; the two are unified (McNeill, Cassell & McCullough, 1994). In each case, the meaningful linkage of gesture and language resists division.

3.2. Growth points as psychological predicates and the context of speaking

A thread that runs through all examples of GPs is the following: GPs inferred from speech-gesture synchronizations are the elements in the immediate context that are presented as

newsworthy, they are points of departure from preceding discourse, what Vygotsky called *psychological predicates* (as opposed to grammatical predicates).

The concept of a psychological predicate illuminates the theoretical link between the GP and the context of speaking. Defining a psychological predicate (and hence a GP) requires reference to the context; this is because the psychological predicate and its context are mutually defining. The GP:

1. marks a significant departure in the immediate context; and

2. implies this context as a background.

We have in this relationship the seeds for a model of realtime utterance generation and coherent text formation.

Regarding the GP as a psychological predicate suggests a mechanism of GP formation in which differentiation of a focus from a background plays an essential part. Such differentiation is validated by a very close temporal connection of gesture strokes with the peak of acoustic output in speech. Nobe has documented this connection instrumentally: "The robust synchrony between gesture strokes and the peaks of acoustic aspects suggests that the information the gesture stroke carries has an intrinsic relationship with the accompanying speech information prominently pronounced with these peaks. The manifestation of the salient information seems to be realized through the synchronization of these two modalities" (Nobe, 1996, p. 35).

Regarding the GP as a psychological predicate also clarifies the sense in which we use the term 'context'. This term has a host of meanings (cf. Duranti & Goodwin, 1992), but for our purposes 'context' is the background from which a psychological predicate is differentiated. This background indexes and is constrained by external conditions, both social and material, but the background is also under the control of the speaker; it is a *mental construction*, part of the speaker's effort to construct a meaning. The speaker shapes the background in a certain way, in order to make possible the intended significant contrast within it. Background and contrast are both necessary and are constructed together.

3.3. Growth points in thinking-for-speaking

Though seemingly antithetical, image and language category are equally indispensable to thinking-for-speaking. The opposition between them is the key that unlocks speech. As image and language interact they are able to influence one another — the "continual movement back and forth" of which Vygotsky spoke in his evocation of the dialectic of language and thought. This mutual influence enables language to influence imagery and imagery to influence language, as the utterance unfolds in real-time. Speech is not a translation of one medium, image, into another, language. To grasp it, we should not replace the minimal unit possessing antithetical poles with a succession of single poles. In keeping with Vygotsky's conception of inner speech the GP, with its dual imagistic-categorial nature, is the mediating link between individual cognition and the language system.

3.4. Unpacking growth points

The dialectic continues during what Werner & Kaplan (1963) termed 'microgenesis' and we will call 'unpacking'. The surface utterance works out the implications of the GP and finds a grammatical framework in which to fit it. Thought undergoes continuous change during this process, thus shaping thinking while speaking.

Linguistic categorization may select only some aspects of the image for classification. The gesture, being global and, especially, being synthetic, is likely to contain information pertaining to more than one component of meaning. The segment(s) of speech with which the gesture stroke synchronizes, however, need not categorize all this information. Since the gesture in (1) displayed two hands releasing or thrusting down, it could logically have been categorized with "drops". However, the stroke of the gesture was withheld during this verb, even though the hands were in position to perform it. In GP terms, the core concept excluded the act of dropping, which was an action by the character, Tweety. This fits with the narrative goal, which was to explain how the bowling ball got inside Sylvester and flushed him out of the pipe. In this logic, the emphasis was on the bowling ball and the key transformation was attributed to it. "Drops" therefore, the action not of the ball but of Tweety, was not part of the GP, and the gesture stroke, accordingly, excluded it. To identify the gesture as "drops" would have been meaningful had

What Tweety Was Doing been the significant opposition, rather than What The Bowling Ball Did.[6]

What processes of unpacking resulted in the verb, "drops", then? This latter component characterized Tweety's action, not the bowling ball's. Thinking subsequently shifted to categorize the action of this agent and supplied a verb referentially motivated by it. According to this explication, a verb is not the only way to anchor a sentence. In this example the verb was essentially *derived* from the GP's complex of meanings. In the course of unpacking the GP, thinking shifted and acquired an agentive cast. In such a case the verb, though uttered first, would arise in the sentence generation process after the GP itself. The dynamics of thinking-for-speaking during unpacking thus highlight the distinction between action by the agent and the other object's path, a distinction appropriate in the discourse context of this GP.[7]

3.5. Linguistic relativity

This distinction may come easily to thinking that is being formulated in English and similarly organized languages, in which path and agent are separably encoded meaning components, but it is not easily achieved in other languages; Georgian for example. In one narration in this language, a downward gesture stroke corresponding to (1) was categorized by a verb that includes both path and agentivity content (*chagdebs*, 'throws-down').[8] In other words, Georgian does not make available a path word outside the verb to categorize just this feature. The image inevitably is categorized for agentivity as well. The Georgian speaker's thinking-for-speaking would thus differ in this respect from the English speaker's, though the imagery was effectively the same.

3.6. Summary of the growth point

To sum up, a GP is neither word nor image. It is thinking in global imagery and linguistic categories simultaneously. Its essential feature is a dialectic of these forms of thinking, and it gives rise to speech and gesture through their collaboration ('convergence' — Kita, this volume). Speech-gesture synchrony is therefore explained genetically, as an inevitable consequence of how the idea unit itself took form and its resistance to interruption during unpacking. Speech-

gesture synchrony could not be otherwise with an initial organizing impulse of this sort. Thinking, according to this hypothesis, is both global *and* segmented, idiosyncratic *and* linguistically patterned. The implied model of language production is therefore not G (imagery)→L; that is, language is not a translation of imagery. Nor is it L→G, meaning that the gesture depends "sequentially and organizationally" on language.[9]

4. MOTION IN THREE LANGUAGES

With the GP framework now described, we will illustrate its application in three languages. We shall search for differences in thinking-for-speaking as embodied in GPs in English, Spanish and Chinese (we have already mentioned Georgian). This approaches, via observable speech and gesture, the experience of having different forms of thinking in different languages.

4.1. Motion events

We focus on a particular semantic domain, the motion event. This domain offers the advantage of allowing us to borrow much current linguistic and psycholinguistic analysis. In particular, Talmy's (1985, 1991) motion event componential analysis provides a cross-linguistic schema in which motion is analyzed into a set of semantic components, and languages are compared and grouped according to how they package these into linguistic forms (see also Aske, 1989; Slobin, 1987; Choi & Bowerman, 1991). According to Talmy, a prototypical motion event expression has these components, among others:

• A moving object, called the 'figure', as in "drops *it* down", where the "it" indexes the bowling ball, the object in motion;

• A reference object, called the 'ground', as in "drops it down *the drainpipe*", where the downward trajectory occurs in relation to a non-moving object, the drainpipe;

• A trajectory, or 'path', as in "drops it *down*", where the bowling ball moves on a downward trajectory;

• A 'manner', as in "it *rolls* down", where the motion is performed in a certain way.

According to Talmy, each given language has a characteristic way of packaging such motion event components. The languages we discuss here fall into two classes that Talmy has classified as 'satellite-framed' and 'verb-framed'. The category depends on how the path component of the motion is packaged. English is satellite-framed, meaning that path is coded in a so-called 'satellite' — i.e., an adjunct to the verb, like *down*. Path is coded outside the main verb. Spanish is verb-framed in that path information is bundled into the verb itself. Talmy (1985) classifies Chinese as also satellite-framed, like English, although other writers have placed it on a continuum somewhere between English and Spanish (Slobin & Hoiting, 1994).

Manner and how it is presented is a second important difference between the verb- and satellite-framed types. In contrast to path, manner in a satellite-framed language is encoded in the main verb and indeed both English and Chinese have rich lexicons of manner verbs (e.g., distinctions of bipedal locomotion — *walk*, *run*, *stroll*, etc.). Manner is prepackaged in many English and Chinese verbs.

Because of this relentless presence of manner, English speakers appear sometimes to go out of their way to avoid it. We find that deictic verbs like *go* and *come* have a use in English discourse as a way of avoiding manner, when including it may be an undesirable over-specification in the motion event description; for example, saying "he comes out the pipe" when "he rolls out the pipe" would have been referentially appropriate but might seem an over-specification in the context (where only the fact of motion was significant). We will show that gesture provides another means by which English speakers exclude manner.

With Spanish, in contrast, path is bundled into the verb while manner is often introduced constructionally outside the verb, in a gerund, a separate phrase, or clause. The gerund is a frequent pattern, illustrated by *sale volando* 'exits flying', in which path is in the main verb and manner is in a gerund. Slobin (1996) has discovered an effect of this typological difference — novels written originally in English lose as much as half of their manner coloration in Spanish language translations, presumably because including manner is rhetorically cumbersome. As we will show, a somewhat more complex picture emerges when gestures are considered.

4.2. Motion event gestures in English and Spanish

   4.2.1. Gestural manner. As mentioned above, there is sometimes an embarrassment of riches in English with respect to manner, and the problem is how to downplay it. We present here a snapshot of how gesture can downplay manner, even when the speaker employs a manner verb in speech. Spanish speakers are an interesting contrast case, for they do not face this same problem.

   In English we see two alternating patterns. In one, there is a focus on manner through gesture; in the other there is use of gesture to downplay manner. The following examples illustrate these patterns with the verb *rolls*:

(2)   [but it **rolls**] him out

    Hand wiggles: manner information.

(3)   [and he rolls ... **down** the drain spout]

    Hand plunges straight down: path information only.

In (2), which is from one speaker, a wiggling-hand gesture synchronized with the "rolls" in the utterance. We infer that the GP of this utterance consisted of manner imagery (here shown with gestural 'agitation') categorized as rolling.   In (3), from another speaker, the gesture lacked manner and did not synchronize with "rolls" at all.   Instead, the gesture skipped the verb and synchronized with path, "down", and ground, "the drainspout" (via the post-stroke hold).

   In other words, even when the same verb occurs, gestures can differ. From this we infer different GPs and thinking-for-speaking.   In one pattern, the core idea includes manner, whereas in the other it does not and the manner in the verb is made to recede in the face of a gesture without manner that may be synchronized with other content, as in (3). We might suspect that these differences of focus occur within contrasting immediate contexts — in (2) a context in which the rolling manner of the ball is a significant opposition *versus* a context in (3) in which manner is not treated as significant. There is some support for this prediction.   The speaker of (2) was focusing on the motion of the bowling ball, while the speaker of (3) focused on

Sylvester. The framework of significant oppositions was therefore different and the gesture placements (and inferred GPs) shifted accordingly.[10]

In Spanish, we find the opposite situation, cases where gestural manner appears without a linguistically encoded anchor and has the potential to appear at several places in a motion event description, something like a manner 'fog.' A manner fog is defined as manner in gesture that is not also coded in speech. Although Spanish speakers often omit manner from their speech, manner is abundant in their gestures and combines with other linguistic categories, typically path (verb) and/or ground (nominal phrase).

(4.1) e entonces busca la ma[ner**a (silent pause)**][11]

and so he looks for the way

Gesture depicts the shape of the pipe: the ground

(.2) [de e**ntra**][r / / **se met**][**e por el**]

to enter REFL goes-into through the

Both hands rock and rise simultaneously: manner and path (left hand only

through "mete")

(.3) [de**sague** _/ /_] [/ / si?]

drainpipe...yes?

Right hand continues to rise with rocking motion: path + manner.

(.4) [de**sague entra** /][12]

drainpipe, enters

Both hands briefly in palm-down position (clambering paws) and then rise with chop-like

motion: path + manner.

Gestural manner (either a climbing or a winding, 'corkscrewing' motion) was articulated throughout this description.[13] Each GP in this bit of discourse could embody manner, but this manner, unlike that in the English example in (2), was not categorized as manner by any linguistic unit. Rather, manner imagery was categorized in non-manner ways — path: "mete" 'goes-into' in (4.2), ground: "desague" 'drainpipe' in (4.3), and path again: "entra" 'enters' in

(4.4). GPs thus brought manner in gesture into the Spanish categorial system through routes other than manner as a category itself. The question is how manner is categorized in Spanish when there is no clear focus for manner given in the verb system. These gestsural manner combinations with Spanish path or ground are GPs, no less than GPs are with manner verbs in English. In Spanish, we say, GPs with manner are categorized and enter the linguistic system as path and/or ground. That manner in gesture goes with path and/or ground is an empirical observation that may say something about the interrelations of motion event concepts themselves, especially manner and path/ground.

Thinking in relation to manner thus can move in opposite directions in English and Spanish. In English, gestural manner focuses at a specific point (the verb) if it is part of a core idea; otherwise it is omitted and the gesture stroke can skip the verb, thus downplaying the manner component. Alternatively, speech finds its way to a non-manner-encoding expression, such as *comes out of*. In Spanish, gestural manner, far from appearing only when it is categorized as manner, can appear in the absence of spoken manner and be categorized instead as a kind of path (4.2 and 4.4) or a kind of ground (4.3).

We interpret this difference between Spanish and English as arising from thinking-for-speaking adaptations. In English, gesture and verb jointly highlight manner when it is part of the speaker's focus. When manner is not in focus, gesture does not encode it, and need not synchronize with a manner verb, even if one is present. Thus gesture modulates the lexical system of English, and while it may include manner it also may exclude it. Spanish, with its non-obligatory manner, does not require modulation. Manner appears in Spanish speech presumably only when it is a focused component and it is often omitted even when it is potentially significant. Finding a way to include manner is the challenge in Spanish. Gesture, again, adds to thinking-for-speaking but in the opposite direction from English. This analysis suggests that differences in the dynamics of thinking-for-speaking can be traced to typological differences between languages.

4.3. Motion event gestures in English and Chinese

Talmy (1985) includes Chinese in the class of satellite-framed languages. Like English, it has a large lexicon of manner verbs and high-frequency constructions that parcel out the components of motion events in ways similar to English. Path, for example, is typically not encoded in the verb but expressed instead in particles associated with the verb. In the gestures of Chinese speakers we find no cases of manner 'fogs', as in Spanish. Thus gestural evidence, too, aligns Chinese and English within this framework.

4.3.1. Speech-gesture framing. In Chinese, however, we find a language-specific pattern of gesture and possibly of thinking-for-speaking. The hallmark of this Chinese pattern is a gesture that occurs earlier in the temporal sequence of speech than the timing facts of English and Spanish would lead us to expect. This suggests a difference in GP possibilities. An example is the following:

(5)   lao tai-tai [na   -ge   da **bang hao**]-xiang gei   ta   da-xia

     old lady hold CLASSIFIER   big stick seem   CAUSE   him   hit-down

                                               verb-satellite

     'The old lady apparently knocked him down with a big stick'

The gesture (a downward blow) that accompanied the spoken reference to the stick was expressively redundant with the verb and satellite, "da-xia" 'hit-down.' As the speaker said "da bang" 'big stick', she performed a downward blow gesture. Her hand then promptly relaxed and went to the rest position well before the verb phrase emerged in speech. This timing pattern is often found in our Chinese data. It presents a quite different picture from the sentential predicate (verb-) focused gestures we see with English and Spanish speakers. It is as if the gesture shifts forward in the surface speech stream, in the direction of the utterance-initial position characteristic of topic statements in Chinese speech. We do not consider such speech-gesture pairings to be 'errors' of synchrony; there nothing random about them. We find, in every such case, an idea unit with clear semantic coherence. In terms of thinking-for-speaking, we interpret this pattern of synchrony as evidence of imagery that forms with speech idea units based on framing constraints: units that specify upcoming domains of reference, before they are

articulated in speech. In English and Spanish, in contrast, the tendency is to think in terms of the kinds of event transformations embodied in the grammatical predicates of subject-predicate constructions. Chinese shows the latter pattern in many instances as well, but adds this further pattern that may resemble topicalization in its framing effects.[14]

Chinese is what Li and Thompson (1976, 1981) have termed a 'topic prominent' language. English and Spanish, in contrast, are 'subject prominent'. Utterances in the latter are founded on subject-predicate relations. In line with this typological distinction, we find cases like (5), in which gesture provides one element and speech another element to jointly create something analogous to a topic frame.[15] Again, therefore, we see the possible impact of language type on thinking in that language.

4.3.2. English in contrast. The Chinese-specific character of this use of imagery for thinking on the topic-like frame level becomes clear when Chinese is compared to English. In English too, a gesture depicting an event yet to be expressed in speech occasionally gets ahead of speech in production. But the precocious imagery appears to be regarded by the speaker as an error, something to be repaired. In example (6.1), a gesture that shows the result of an action synchronizes with speech describing the cause, a semantically appropriate pairing:

(6.1)   [so it **hits** <u>him on</u> **the hea**][<u>d</u>

(.2)   <u>and he winds up</u> **rolling down the stre**]et

The two gestures in (6.1) depicted Sylvester moving down the street, an event not described until (6.2). In (6.1) they synchronized with a description of the initiating condition for the gestured event; the pairing therefore is similar to that in the Chinese example. The difference between the languages is apparent at the next step. Unlike the Chinese speaker, this English speaker held and then repeated the gesture in a larger, more definite way when the target linguistic segments emerged in speech (6.2). Thus, an important difference between Chinese and English thinking-for-speaking dynamics is exposed by this contrast between seemingly similar speech-gesture combinations. No hold or repetition occurred in the Chinese example. The subsequent enhanced repeat in the English example indicates the relevance of the gesture to the predicate. This is what

makes it look like a gestural 'repair'.   In other words, she retained the imagery at (6.1) for the GP of (6.2). She did not use it, as did the Chinese speakers, as a self-contained framing unit.

       4.3.3. 'Enslaved' to predication. Gestural evidence that links the typological distinction of topic- vs. subject-prominence with different patterns of thinking-for-speaking indicates that Chinese speakers are able to create a topicalizing frame with conjoined gestures and speech, something that appears to be beyond the ken of English (and, presumably, Spanish) speakers. The packaging of motion information in English (and, presumably, Spanish) is encased within the grammatical structure of the sentiential predicate to a degree not found in Chinese. In English, given one component of a motion event, others are bound to be included in a structure more or less preordained, particularly by the predication structure and its linkage to other parts of the sentence. We can think of English and Spanish as in this sense 'enslaved' to predication because they need to find machinery (grammatical, morphological, gestural) to link predicates to their subjects. In the English example above, there is a *linking gesture hold*, which functions as a kind of anaphoric reference. In terms of the GP hypothesis, the expatiation of GPs in English and Spanish is satisfied by structures typically organized as sentiential predicates. This constituent is thus at the heart of thinking-for-speaking in these languages. In Chinese, in contrast, the "da-xia" 'hit-down' verb-satellite assembly in (5) was dissociated from gesture and intonationally de-stressed, as if its role in speech was largely formal.   Thinking-for-speaking in Chinese therefore seems less tied to specific predicate-based constructions.

4.4. Summary of motion event expression

       We have seen evidence of the following:

       1. The GP is a unit of analysis applicable in all languages.

       2. The minimal GP unit of thinking is irreducibly imagery and a linguistic category.

       3. Describing the same motion events, languages encourage different forms of thinking. English and Spanish (as well as Georgian) are predicative in their focus, but thinking differs in how motion event semantics are focused. Chinese induces thinking in which the focus is a frame

for other information. Observations thus show an effect of linguistic organization on thinking on two levels — predicative and discourse — and different patterns on both.

4. As a model of thinking-for-speaking, the GP embodies thinking that exists because of the linguistic code. Thinking emerges via the GP with language categories built in.

5. EMBODIED COGNITION AND BEING

Why do we perform gestures at all? In what sense are gesture and speech embodied cognition? What functions do gestures serve for thinking-for-speaking? In the cases we have described, we can see that the relationship of gestures to speech is shaped by language. Researchers have asked if gesture is motivated mainly or entirely by a desire to convey narrative- or discourse-pragmatic level information that is less than adequately conveyed by the linguistic code. If this is the role of gesture, certain phenomena must be carefully explained. Gestures occur, famously, in the absence of a face-to-face audience, on the phone for example[16]. Further, they often do not reflect the deliberateness of speech. They do not occur invariably, so there is some selection principle. As well, when they occur they may be more or less elaborated as kinesic performances.

Several proposals can be mentioned to explain the occurrence of gestures, among them that they are the remnants of an earlier stage of language evolution (Armstrong, et al., 1995; Donald, 1991), and that they have their own communicative effects.   Speech-gesture patterning is undoubtedly heterogeneous in origin; it is shaped by culture, possibly by evolution, and includes social-interactional as well as individual factors. In developing a theory of language production centered on the GP as a unit of analysis, we wish to address the implications of the idea that gesture, the actual kinesic event itself, is a dimension of thinking.

Gestures are *material carriers* of thinking — a phrase used somewhere by Vygotsky. Note that in the GP model, speech is equally a dimension of thinking. The same sense of a material carrier is thus invoked. Speech and gesture are best thought of, we argue, not as the packaged communicative outputs of a separate internal production process but, rather, as the joint embodiment of that process itself. One approach to the material carrier concept is found in

Werner & Kaplan (1963), who wrote of the 'organismic' foundations of symbolization; of our capacity to represent the world symbolically by capturing it in a kind of imitation carried out in bodily movements made possible by "... this *transcendence of expressive qualities*, that is, their amenability to materialization in disparate things and happenings ..." (p. 21). The development of an individual child in this view is, in part, a process of adding semiotic distance between movement and the expressive qualities it can have: "The act of denotative reference does not merely, or mainly, operate with *already formed* expressive similarities between entities. Through its productive nature, it brings to the fore latent expressive qualities in both vehicular material and referent that will allow the *establishment of semantic correspondence* between the two entities. It is precisely this productive nature of the denotative act that renders possible a symbolic relation between any entity and another." (pp. 21-22, in both passages, emphasis in the original). We see this creation of meaning in gesture.

A more radical extension of the material carrier concept carries it past the realms of representations. This to give the concept an interpretation on the level of cognitive being. To the speaker, gesture and speech are not only 'messages' or communications, but are a way of cognitively existing, of cognitively being, at the moment of speaking. By performing the gesture, the core idea is brought into concrete existence and becomes part of the speaker's own existence at that moment. The Heideggerian echo in this statement is intended. Gestures (and words, etc., as well) are themselves thinking in one of its many forms — not only expressions *but thought, i.e., cognitive being, itself.* The speaker who creates a gesture of Sylvester rising up fused with the pipe's hollowness is, according to this interpretation, embodying thought in gesture, and this action — thought in action — was part of the person's being cognitively at that moment. To make a gesture, from this perspective, is to bring thought into existence on a concrete plane, just as writing out a word can have a similar effect. The greater the felt departure of the thought from the immediate context, the more likely its materialization in a gesture, because of this contribution to being. Thus gestures are more or less elaborated depending on the importance of material realization to the existence of the thought. Such a correlation of

gesture elaboration with less continuous/predictable references in speech has been observed (McNeill, 1992, p. 211).

There are, however, deep and hitherto unexplored issues here, and possibly some contradictions. If to the speaker the gesture and linguistic form are themselves forms of being cognitively, there would seem to be no room in this process for the presentation of symbols; the signifier-signified distinction that constitutes semiosis is lacking. A semiotic relation appears when an observer is taken into account — a listener who participates or a coder who looks at the communicating individual. Dreyfus (1994), in his invaluable exposition of Heidegger, explains Heidegger's treatment of symbols in a way that suggests a rapprochement. To cope with signs is not to cope just with them but with the whole interconnected pattern of activity in which they are embedded (this still has the viewpoint of a listener/observer; from the speaker's viewpoint, we should say that producing a sign *carries the speaker* into a 'whole interconnected activity'). Heidegger, according to Dreyfus, says that signs point out the context of a shared practical activity — and this is the key to the rapprochement. To have your thoughts come to exist in the form of signs is to cause them to exist in a context of shared practical activities. A sign signifies only for those who 'dwell' in that context. This we can recognize is a recipe for the GP: sign and context are inseparable and this context must be dwelled in. This brings the GP and the social interactive context together as joint inhabitants of the context (and it is the speaker who always must be the one dwelling there the best). The communication process is then getting the Other to dwell there on her own. In this way the GP model can be seen to map 'external' interactive contexts into internal units of functioning, a convergence of this mode of theorizing onto Vygotsky's model of two planes, the interpsychic and intrapsychic.[17]

In a GP the gesture, as one dimension of a material carrier, is as much a part of the linguistic process as are the familiar linguistic objects of words, phrases and clauses. The gesture adds substance to the speaker's cognitive being. When, by gesture, language is extended, narrowed, or adapted to exploit a linguistic feature, cognitive being itself is changed.

6. CONCLUSION

We have argued in this paper that speakers of different languages create language-specific modes of thinking-for-speaking. Gesture contributes material carriers to thinking-for-speaking and these take different forms in different languages. Speech and gesture together can be conceptualized as bringing thinking into existence as modes of cognitive being. This concept explains the occurrence of gestures, and explains why they are more frequent and more elaborate in contexts where the departure of the meaning is felt to be greater.

<u>REFERENCES</u>

Armstrong, D.F., Stokoe, W.C., & Wilcox, S.E. (1995). <u>Gesture and the Nature of Language</u>. Cambridge: Cambridge University Press.

Aske, J. (1989). Path predicates in English and Spanish: A closer look. In <u>Proceedings of the 15th Annual Meeting of the Berkeley Linguistics Society</u>. Berkeley, CA: Berkeley Linguistics Society.

Bolinger, D. 1986. <u>Intonation and its parts</u>. Stanford: Stanford University Press.

Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C.N. Li (ed.), <u>Subject and Topic</u> (pp. 25-55). New York: Academic Press.

Choi, S. & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. <u>Cognition</u>, **41**, 83-121.

Cohen, A.A. (1977). The communicative functions of hand illustrators. <u>Journal of Communication</u>, **27**, 54-63.

Donald, M. (1991). <u>Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition</u>. Cambridge, MA: Harvard University Press.

Dreyfus, H.L. (1994). <u>Being-in-the-World: A Commentary on Heidegger's Being and Time, Division I</u>. Cambridge, MA: MIT Press.

Duncan, S.D. (1996). Grammatical Form and 'Thinking-for-Speaking' in Chinese and English: An Analysis Basedon Speech-Accompanying Gestures. Ph.D. dissertation, Department of Psychology (Cognition and Communication) and Department of Linguistics, University of Chicago.

Duncan, S.D. (in progress). Grammaticalization of the verb "na" in spoken Mandarin Chinese. Max Planck Institute for Psycholinguistics (CARG), Nijmegen.

Duncan, S.D., McNeill, D., & McCullough, K.-E. (1995). How to transcribe the invisible — and what we see. In D.C. O'Connell , S. Kowal, & R. Posner, (eds.), <u>KODICAS/CODE</u> (special issue on Zeichen für Zeit: Zur Notation und Transkription von Bewegungsabläufen), **18**, 75-94.

Duranti, A. & Goodwin, C. (1992). <u>Rethinking Context: Language as an Interactive Phenomenon</u>. Cambridge: Cambridge University Press.

Ekman, P. & Friesen,

Fox, B. (1995). On the embodied nature of grammar: Embodied being-in-the-world. Paper presented at the International Conference on Functional Approaches to Grammar. University of New Mexico (July).

Harré, R. (1986). Mind as social formation. In J. Margolis, M. Krausz & R.M. Burian (eds.). <u>Rationality, Relativism and the Human Sciences</u>!(pp. 91-106). Dordrecht: Martinus Nijoff.

Harris, R. & Taylor, T.J. <u>Landmarks in Linguistic Thought: The Western Tradition from Socrates to Saussure</u>. London: Routledge, 1989.

Kendon, A. (1972). Some relationships between body motion and speech: An analysis of an example. In A. Siegman & B. Pope (eds.), <u>Studies in Dyadic Communication</u> (pp. 177-210). Elmsford, NY: Pergamon.

Kendon, A. (1977). Studies in the Behavior of Social Interaction. Bloomington, IN: Indiana University Press.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M.R. Key (ed.), <u>The Relationship of Verbal and Nonverbal Communication</u> (pp. 207-227). The Hague: Mouton.

Kendon, A. (1995). Gestures as illocutionary and discourse structure markers in Southern Italian conversation. <u>Journal of Pragmatics</u>, **23**, 247-279.

Kendon, A. (this volume).

Kita, S. (1990). The temporal relationship between gesture and speech: A study of Japanese-English bilinguals. Unpublished MS thesis. Department of Psychology. The University of Chicago.

Kita, S. (1993). <u>Language and Thought Interface: A Study of Spontaneous Gestures and Japanese Mimetics</u>. Ph.D. dissertation, Department of Psychology (Cognition and Communication) and Department of Linguistics, University of Chicago.

Kita, S. & Wang, X. (1992). Co-occurrence of iconic gestures and "onomatopoeias" in Japanese and Chinese narrations. Paper presented at the 5th annual convention of the American Psychological Society (Chicago, June).

Levinson, S. (in press). The body in space: cultural differences in the use of body-schema for spatial thinking and gesture. In G.Lewis & F.Sigaud (eds), <u>Culture and the Uses of the Body</u>. Oxford Univ. Press.

Li, C. N. & Thompson, S. A. (1976). Subject and topic: A new typology of language. In C.N. Li (ed.), <u>Subject and Topic</u>. New York: Academic Press.

Li, C. N. & Thompson, S.A. (1981). <u>Mandarin Chinese: A functional Reference Grammar</u>. Berkeley: University of California Press.

Lucy, J.A. (1987). <u>Grammatical Categories and Cognition: An Historical, Theoretical, and Empirical Re-evaluation of the Linguistic Relativity Hypothesis</u>. Cambridge: Cambridge University Press.

Lucy, J.A. (1992). <u>Grammatical Categories and Cognition: A Case Study of the Linguistic Relativity Hypothesis</u>. Cambridge: Cambridge University Press.

Lucy, J.A. & Wertsch, J.V. (1987). Vygotsky and Whorf: A Comparative Analysis. In M. Hickmann (ed.), <u>Social and Functional Approaches to Language and Thought</u> (pp. 67-86). Orlando, FL: Academic Press.

Mayberry, R.I. &Jaques, J. (this volume). The disruption of gesture by stuttering: Insights into the nature of the gesture-speech integration.

McNeill, D. (1992). <u>Hand and Mind: What Gestures Reveal about Thought</u>. Chicago: University of Chicago Press.

McNeill, D., Cassell, J., & McCullough, K.-E. (1994). Communicative effects of speech-mismatched gestures. <u>Research on Language and Social Action.</u> **27**, 223-237.

Morris, D., Collett, P., Marsh, P. & O'Saughnessy, M. (1979). <u>Gestures: Their Origins and Distribution</u>. New York: Stein & Day.

Özyürek, A. (forthcoming). Social Space as the Common Indexical Ground of Gesture and Language Use. Ph.D. dissertation, Department of Psychology (Developmental Psychology) and Department of Linguistics, University of Chicago.

Özyürek, A. (this volume).

Saussure, F. de (1916). Course in General Linguistics. New York: Philosophical
     Library.
Schegloff, E.A. (1984). On some gestures' relation to talk. In J. Maxwell & J.
     Heritage (eds.), Studies in Conversation Analysis (pp. 266-296).
     Cambridge: Cambridge University Press.
Slobin, D. (1987). Thinking for speaking. In J. Aske, N. Beery, L. Michaelis, & H.
     Filip (eds)., Papers from the 13th Annual Meeting of the Berkeley Linguistics
     Society. Berkeley, CA: BLS.
Slobin, D. (1996). Two ways to travel: Verbs of motion in English and Spanish. In
     M. Shibatani and S. A. Thompson (eds.), Grammatical Constructions: Their
     Form and Meaning. Oxford: Oxford University Press.
Slobin, D. & Hoiting, N. (1994). Reference to movement in spoken and signed
     languages. In Proceedings of the 20th Annual Meeting of the Berkeley
     Linguistics Society. Berkeley, CA: Berkeley Linguistics Society.
Streeck, J. (1995). The interactive production of visual communication.
     Manuscript in preparation, University of Texas, Austin (1995).
Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In
     T. Shopen (ed.), Language Typology and Syntactic Description. Vol. III:
     Grammatical Categories and the Lexicon (pp. 57-149). Cambridge:
     Cambridge University Press.
Talmy, L. (1991). Path to realization: A typology of event conflation. In
     Proceedings of the 17th Annual Meeting of the Berkeley Linguistics Society.
     Berkeley, CA: Berkeley Linguistics Society.
Vygotsky, L.S. (1987). Thinking and Speech. In Robert W. Rieber & Aaron S.
     Carton (eds.), Norris Minick, (trans.), 39-285.
Werner, H. & Kaplan, B. (1963). Symbol Formation: An Organismic-
     Developmental Approach to Language and the Expression of Thought. New
     York: Wiley.
Whorf, B.L. (1956). Language, Thought, and Reality: Selected Writings of
     Benjamin Lee Whorf, J.B. Carroll (ed.). Cambridge, MA: MIT Press and
     New York: John Wiley.

NOTES

––––––––––––––––––––––––

[2] The expression, 'thinking-*for*-speaking' suggests a temporal sequence: thinking first, speaking second. We posit instead an extended process of thinking-*while*-speaking, but keep the thinking-*for*-speaking formulation to maintain continuity with Slobin and his writings, and to capture the sense of an adaptive function also conveyed by use of *for*, with the caveat that we do not mean by this a thinking→speaking sequence.

[3] 'Idiosyncratic' in this use does not mean unique or bizarre. Two speakers can non-uniquely perform similar non-bizarre (even mundane) gestures, and both be 'idiosyncratic', that is, not meet external standards of well-formedness.

[4] All examples in this paper are from narrations by adult speakers retelling a 7 minute animated color cartoon to a naive listener. Neither speaker nor listener was aware that the speaker's gestures were of interest. For details of the method and transcription, see McNeill (1992) and Duncan, et al. (1995).

[5] Brackets show when the hands were in motion, boldface marks the gesture stroke phase, the phase of 'effort' that bears the semantic content of the gesture, and double underlining marks a hold — the hands held in place in midair — which in this case included both a 'pre-stroke' and a 'post-stroke' hold (Kita, 1990).

[6] To describe a gesture as displaying components of motion at all is, in a sense, an oxymoron, in that takes linguistically segmentable meanings and attributes them to a quite different kind of semiotic entity. The holistic stroke in "drops it down" was seamlessly all these motion components at once: 'pathgroundfiguremanner'. The synchronized speech brought into segmented existence the first three components of this amalgam.

[7] See McNeill, this volume, for analysis of contexts in GP formation

[8] We are grateful to Kevin Tuite for this example.

[9] To use a phrase of Shegloff's (1984). The GP can also be compared to his concept of a 'projection space'. While evidently related, projection space and GP are not identical. They apply to different empirical domains — conversational interaction, and the microgenesis of individual speech and thought, respectively.   Although addressing different domains, we can answer Schegloff's question, what is 'in play'?   It is *the very GP itself* that is 'in play' (not a word or lexical affiliate, therefore, but a linguistically categorized image).

[10] Example (2) was at the end of a series of references to the bowling ball where it and what it was doing would have been highlighted.

and he drops a [bowl]ing ball [in**to** <u>the rain spout</u>]

[and it **goes do**wn]

and it* [/] ah*

you [**can't tell if the bow**<u>ling ball</u> /] [is un* /] [is **und**<u>er Sylvester</u>

<u>or</u> **ins**ide of him]

[but it **rolls him out**]*   (=2)

Example (3) appeared in a series that began similarly but then shifted to Sylvester and his path. The shift took place before (3) and continued beyond it, and would have created a context in which the bowling ball and its manner of motion would be downplayed.

[the canary] # [th**row**s*] # [puts a # [bowling] [ball] #

into] # [the **drain** <u>spout as the</u>]

[**cat** <u>is climbing up /and</u>] [it **goes** <u>into his</u>] [mouth] /   (switch to Sylvester)

[and **of course**] # [**into** <u>his</u> stomach] #

[and he rolls # **down** <u>the drain</u> spout]   (=3)

[and [**across**] [the **street**] into [the bowling] alley #]

[11] Kendon (personal communication) believes that when speech halts like this, listeners have a tendency to shift their gaze to the speaker, and this could be such an occasion.

[12] We are grateful to Lisa Miotto, Karl-Erik McCullough, and Gale Grubman-Stam for transcribing and translating this example. The example is from Grubman-Stam's data.

[13] In the grammar of Spanish, the combination of manner and path is confined to paths that do not end with the moving object entering a specifically configured state (Aske, 1989; Slobin & Hoiting, 1994). In Example (4), however, path *does* end in a 'specifically configured' ground state — the drainpipe; yet it appears with gestural manner. The GP, as a dual imagery-language unit, can expand the resources of the language at such points (cf. Kita, this volume).

[14] Chafe (1976) stated our intended sense of topicalization : "What the topics appear to do is limit the applicability of the main predication to a certain restricted domain ... the topic sets a spatial, temporal, or individual framework within which the main predication holds." (p. 50; also quoted by Li and Thompson, 1976).

[15] Gestures may be recruited for topicalization precisely where they can expand the resources of the linguistic system.   Duncan (1996) found that accompanying gesture reflects the particular semantic link between the case-marking grammatical particle *ba* and the nominal it marks. Duncan (in progress) is exploring the possibility that gestures are recruited at points of incipient grammaticalization. Example (5) represents such a point involving the verb *na*, 'to pick up', which often appears as in the example given where speech and gesture jointly specify the framework within which the main predication holds. The verb *na* may have an incipient grammaticized role of marking instrumentality (the big stick mentioned).

[16] Although the number of gestures is reduced when a social presence is lacking, as in speaking to a tape recorder (Cohen, 1977).

[17] The Vygotskian model of two planes, the 'interpsychic' and the 'intrapsychic', helps clarify the relationship of individual cognition to the social context of speaking. In Duranti & Goodwin's (1992) discussion of Vygotsky, however, the focus is exclusively on the interpsychic

plane (Harré, 1986, performs a similar 'intraectomy'). This transmogrification of Vygotsky

removes any basis for considering the relationship of individual cognition to social context.